

概率论与数理统计

本篇笔记内容主要整理自笔者的教材——《概率论与数理统计》（第四版），作者为盛骤、试式干、潘承毅等人，高等教育出版社出版。

一、概率论的基本概念

1. 什么是概率？

描述性定义：随机事件A发生的可能性的度量的大小（非负值），称为事件A发生的概率。

公理化定义：在随机试验的样本空间 Ω 的每一个事件A，都对应一个实数值 $P(A)$ ，如果函数 $P(\cdot)$ 满足下列条件：

- 非负性： $\forall A \in \Omega, P(A) \geq 0$
- 规范性： S 是必然事件，有 $P(S) = 1$ ；
- 可列可加性：设 A_1, A_2, \dots 是两两不相容的事件（即 $i \neq j$ 时， $A_i A_j = \emptyset$ ），有

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n)$$

不相容事件的并的概率 等于 这些事件的概率的和。

2. 古典概型有什么特点？

随机试验的样本空间只包含有限个元素；

随机试验中的每个基本事件发生的可能性都相同。

3. 几何概型有什么特点？

样本空间 Ω 是一个可度量的有界区域；

有无限个基本事件，每个基本事件发生的可能性都一样，即样本点落入 Ω 的某一个可度量区域 S 可能性与 S 的几何度量成正比，而与 S 的位置及形状无关。

4. 什么是条件概率？

在已知事件A发生的情况下事件B发生的概率为条件概率 $P(A|B)$ ，公式有

$$P(B|A) = P(AB)/P(A)$$

5. 什么是全概率公式？

有一些时候事件B的概率不容易直接求，可以通过计算给B在各个条件下 A_i 发生的概率 $P(B|\cdot)$ ，来研究B发生的概率。

$$P(B) = \sum_{i=1}^n P(A_i)P(B|A_i)$$

6. 什么是贝叶斯公式？

$$P(A_i|B) = P(A_i) \frac{P(B|A_i)}{P(B)} = P(A_i) \frac{P(B|A_i)}{\sum_{j=1}^n P(B|A_j)P(A_j)}$$

解释一下“先验”和“后验”的概念（按照课本的思路）

- 1 | P(A)可以通过以往的数据分析得到，可称P(A)是A的先验概率，之所以称为“先验”是因为它不考虑任何B方面的因素。
- 2 | 而在知道B发生的信息之后，来修正A事件发生的概率 $P(A|B)$ ，称为A的后验概率。

通过已知信息B来修正A发生的概率 $P(A|B)$ （即后验概率），可以通过先验概率P(A)以及AB之间的关系 $P(B|A)$ 来研究。

举个例子：假设由多年的统计数据可以知道某种疾病的发病率 $P(A)$ ，有一种检测试剂的准确率为99%，即 $P(B|A)=99\%$ ，同时有 $P(B|\bar{A})=5\%$ 会误报（检测没有病的病人为阳性），可以通过全概率公式计算试剂表现为阳性的概率

$$P(B) = P(A)P(B|A) + P(\bar{A})P(B|\bar{A})。$$

根据这些信息，就可以计算一个病人在这种**试剂检测为阳性的情况下患病**的概率

$$P(A|B) = P(A) \frac{P(B|A)}{P(B)}$$

7. 什么叫做事件相互独立？

$$P(AB) = P(A)P(B)$$

即一个事件的发生，不会影响另一个事件的发生。

二、随机变量及其分布

1. 你知道有哪些常见的离散型随机变量的分布吗？

(0-1) 分布，二项分布，泊松分布。

2. 你说说你对二项分布的理解吗？

只有两种可能结果 A, \bar{A} 的试验称为伯努利试验。二项分布可以理解为进行 n 次独立重复的伯努利试验后, A 发生的次数 (表示为 X) 的概率分布情况。 $X \sim B(n, p)$

3. 你能说说你对泊松分布的理解吗?

泊松分布适合于描述单位时间内随机事件发生的次数的概率分布。例如生活中某医院一天内的急诊病人数, 某地区一段时间内发生的交通事故的次数都可以用泊松分布来建模分析。

PS: 当二项分布的 n 很大, p 很小, $\lambda = np$ 的值适中, 可以用泊松分布来模拟二项分布。

4. 随机变量的分布函数有什么特点?

- $F(x)$ 是不减函数
- $F(-\infty) = 0, F(\infty) = 1$
- $F(x+0) = F(x)$ 即 $F(x)$ 是右连续的。

5. 你知道有哪些常见的连续型变量的分布吗?

均匀分布, 指数分布, 正态分布

6. 什么是指数分布的**无记忆性**?

$$P\{X > s + t | X > s\} = P\{X > t\}$$

举例子来理解: 如果 X 表示一个电子元件的寿命, 且 X 服从指数分布, 那么已知该元件使用了 s 小时接下来元件还能够使用 t 小时的概率, 与该元件从出厂后能使用 t 个小时的概率相同。换句话说, 元件对于它已经使用过的 s 小时没有记忆。

7. 指数分布和泊松分布之间有什么联系?

泊松分布可以理解为: 如果某个事件以固定强度 (λ), 随机且独立地发生, 那么一段时间内发生事件的次数服从的概率分布为泊松分布。

$$P\{X = k\} = \frac{\lambda^k e^{-\lambda}}{k!}$$

而指数分布可以理解为: 相同的独立的随机事件发生的时间间隔服从的分布。

$$F(y) = P\{Y \leq y\} = 1 - e^{-\lambda y} (y \geq 0)$$

注意两条式子的 λ 是一致的, 也就是说如果单位时间内一个事件发生的次数 X 服从参数为 λ 的泊松分布, 那么相邻两个事件发生的时间间隔 Y 服从参数为 λ 的指数分布。

二维随机变量的内容不知道整理什么, 直接跳过了。

四、随机变量的数字特征

1. 什么是期望？什么是方差？

期望是描述随机变量依照概率加权下的平均值。

$$E(X) = \sum_{k=1}^{\infty} x_k p_k$$

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

方差用来描述随机变量与其期望的偏离程度。

$$D(X) = \sum_{k=1}^{\infty} [x_k - E(X)]^2 p_k = E(X^2) - [E(X)]^2$$

2. 什么是切比雪夫不等式？切比雪夫不等式有什么用？

(期望为 μ ，方差为 δ^2)

$$P\{|X - \mu| < \epsilon\} \geq 1 - \frac{\delta^2}{\epsilon^2}$$

在已知期望和方差，却不知道分布函数时，估算随机变量X偏离期望的概率。

3. 什么是协方差？

$$CoV(X, Y) = E\{(X - EX)(Y - EY)\}$$

描述两个随机变量变化程度的一致性。

4. 什么是相关？不相关和独立有什么区别和联系？

两个随机变量XY满足 $CoV(X, Y) = 0$ ，则两个随机变量**不相关**。

独立 \rightarrow 不相关

当XY服从正态分布时，独立 \iff 不相关

五、大数定律及中心极限定理

1. 大数定律

弱大数定理（辛钦大数定理）：一组**独立同分布**的随机变量序列，具有期望 μ ，则随机变量的算数平均值 $\frac{1}{n} \sum_{k=1}^n X_k$ **依概率收敛**于期望 μ 。 $\bar{X} \xrightarrow{P} \mu$

2. 大数定律的推论——伯努利大数定理

f_A 表示 n 次独立重复试验中 A 发生的次数, p 为每次试验中 A 发生的概率, 则

$$\frac{f_A}{n} \xrightarrow{P} p$$

理解: 当 n 足够大时, n 次重复试验后 A 发生的频率, 稳定在概率 p 附近。

这个定理建立起了频率和概率之间的联系, 在生活中, 当试验次数足够大时, 可以用事件的频率来代替事件的概率。

3. 说说中心极限定理

• 独立同分布的中心极限定理

设随机变量 X_1, X_2, \dots, X_n 相互独立, 服从统一分布, 且具有期望 μ , 方差 δ^2 , 则随机变量之和的标准化变量

$$Y_n = \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\delta} \text{ 近似地服从 } N(0, 1)$$

$$\lim_{n \rightarrow \infty} F_n(x) = \Phi(x)$$

• 李雅普诺夫 (Lyapunov定理)

随机变量 X_1, X_2, \dots, X_n 相互独立, 它们具有 $E(X_k) = \mu_k, D(X_k) = \delta_k^2 > 0$, 记 $B_n^2 = \sum_{k=1}^n \delta_k^2$, 则有

$$Z_n = \frac{\sum_{k=1}^n X_k - \sum_{k=1}^n \mu_k}{B_n} \text{ 近似地服从 } N(0, 1)$$

(将上一条定理的随机变量同分布的要求放宽)

生活中很多问题的随机变量可以表示成很多个独立的随机变量之和, 意味着这个随机变量可以用正态分布来近似地描述。

• 棣莫弗-拉普拉斯 (De Moivre-Laplace) 定理:

设随机变量 $\eta_n (n = 1, 2, \dots)$ 服从参数为 $n, p (0 < p < 1)$ 的二项分布, 则对于任意 x , 有

$$\lim_{n \rightarrow \infty} P\left\{ \frac{\eta_n - np}{\sqrt{np(1-p)}} \leq x \right\} = \Phi(x)$$

简单来说, 就是将独立同分布的中心极限定理的分布固定为 (0-1) 分布, 即

$\eta_n = \sum_{k=1}^n X_k$, 代入 $E(X_k) = p, D(X_k) = p(1-p)$ 可证。

理解：正态分布是二项分布的极限分布，当 n 充分大时，可以利用正态分布来估算二项分布的概率。

六、样本及抽样分布

数理统计以概率论为理论基础，数理统计的内容是根据实验或观察得到的数据，来研究随机现象，对研究对象的客观规律作出种种合理的估算和判断。

1. 何为统计量？

统计量是关于样本 $X_1, X_2, X_3, \dots, X_n$ 的 n 元函数，其中函数不包含任何未知参数。

例如，样本均值就是一个统计量($\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$)，表示将各个样本的平均值。

2. 样本方差 $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ 为什么分母是 $n-1$ ，而不是 n ？

因为样本方差 $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ 才是总体方差的无偏估计，而估计量 $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ 不是无偏估计。

更进一步地讲，无偏估计意味着 $E(S^2) = \delta^2$ ，所以由大数定理，取样足够多时， S^2 的均值会依概率收敛到期望，即总体的方差 δ^2 。

3. 统计学的三大分布：

- χ^2 分布
- t分布
- F分布

七、参数估计

1. 什么是点估计？

点估计：总体 X 的分布函数形式已知，但它的一个或多个参数未知，借助于总体 X 的样本来估计总体未知参数值的问题，称为参数的点估计问题。

常见的方法有：矩估计法，最大似然估计法。

2. 什么是无偏性？什么是有效性？

如果一个估计量 $\hat{\theta}$ 的数学期望存在，且等于其估计的参数 θ ，则称为 $\hat{\theta}$ 是 θ 的无偏估计量。

估计量具有**无偏性**可以理解为：用一个估计量反复使用多次计算出来的平均估计值，与真实值的误差几乎可以忽略。（请联系辛钦大数定理）

有效性：比较两个**无偏估计量** $\hat{\theta}_1$ 和 $\hat{\theta}_2$ ，在样本容量 n 相同的情况下，若对于任意 $\theta \in \Theta$ ，有

$$D(\hat{\theta}_1) \leq D(\hat{\theta}_2)$$

且至少对于某一个 $\theta \in \Theta$ ，等号成立，则称 $\hat{\theta}_1$ 较 $\hat{\theta}_2$ **有效**。

3. 什么是相合性？

当 $n \rightarrow \infty$ 时，参数的估计量 $\hat{\theta}$ 依概率收敛于 θ ，则称 $\hat{\theta}$ 为 θ 的**相合估计量**。

4. 什么是区间估计？

人们在估计一个未知量时，有时候希望得到近似值的精确程度，则会估计给出一个范围并包含该参数取真值的可信程度，这种估计称为区间估计，给定的区间称为置信区间。

当未知参数为 θ ，置信区间为 (a, b) 时， $P\{a < \theta < b\} \geq 1 - \alpha$ ，则称区间 (a, b) 的置信水平为 $1 - \alpha$ 。

（注意经常用 α 表示误差，是一个很小的数，所以 $1 - \alpha$ 表示的是可信度）

八、假设检验

1. 假设检验

对总体的某些未知特性，提出关于总体的假设，根据样本的情况对所提出的假设作出是否接受或拒绝的决策。

2. 正态总体均值、方差的检验法（显著性水平为 α ）

见课本P189。主要是看一看，知道统计的三大分布在假设检验中的作用。

PS：后面的章节，包括方差分析及回归分析，bootstrap，随机过程，马尔科夫链等内容，由于没学过就略过了。